Yifeng Ding

yifeng6@illinois.edu
yifeng-ding.com

Research Interests

My research interest lies in Software Engineering and Machine Learning. Specifically, I focus on building large language models to solve software engineering tasks, with a specific interest in improving LLMs' **reasoning and planning** capabilities for **code generation and repair** through **pre-training** [1] and **post-training** [2, 3, 4, 7].

Education

Since 2022 **University of Illinois Urbana-Champaign (UIUC)**,
Ph.D. student in Computer Science,
Advisor: Lingming Zhang

2018–2022 **Tsinghua University**,
B.S. in Software Engineering,
Double Major: B.S. in Business Administration

Publications & Preprints

[1] **Yifeng Ding**, Hantian Ding, Shiqi Wang, Qing Sun, Varun Kumar, and Zijian Wang. "Horizon-Length Prediction: Advancing Fill-in-the-Middle Capabilities for Code Generation with Lookahead Planning". *arXiv; NeurIPS 2024 System 2 Reasoning Workshop*. [preprint].

[2] **Yifeng Ding**, Jiawei Liu, Yuxiang Wei, Terry Yue Zhuo, and Lingming Zhang. "XFT: Unlocking the Power of Code Instruction Tuning by Simply Merging Upcycled Mixture-of-Experts". *62nd Annual Meeting of the Association for Computational Linguistics (**ACL'24**)*. [paper] [code].

[3] Yuxiang Wei, Federico Cassano, Jiawei Liu, **Yifeng Ding**, Naman Jain, Zachary Mueller, Harm de Vries, Leandro Von Werra, Arjun Guha, and Lingming Zhang. "SelfCodeAlign: Self-Alignment for Code Generation". *Thirty-eighth Conference on Neural Information Processing Systems (**NeurIPS'24**)*. [paper] [code] [model] [dataset].

[4] Yuxiang Wei, Zhe Wang, Jiawei Liu, **Yifeng Ding**, and Lingming Zhang. "Magicoder: Source Code Is All You Need". *Forty-first International Conference on Machine Learning (**ICML'24**)*. [paper] [code] [model] [dataset].

[5] Jiawei Liu, Songrun Xie, Junhao Wang, Yuxiang Wei, **Yifeng Ding**, and Lingming Zhang. "Evaluating Language Models for Efficient Code Generation". *First Conference on Language Modeling (**COLM'24**)*. [paper] [code].

[6] Jiawei Liu, Jia Le Tian, Vijay Daita, Yuxiang Wei, **Yifeng Ding**, Yuhan Katherine Wang, Jun Yang, and Lingming Zhang. "RepoQA: Evaluating Long Context Code Understanding". *First Workshop on Long-Context Foundation Models @ ICML'24*. [paper] [code].

[7] Chunqiu Steven Xia, **Yifeng Ding**, and Lingming Zhang. "The Plastic Surgery Hypothesis in the Era of Large Language Models". *38th IEEE/ACM International Conference on Automated Software Engineering (**ASE'23**)*. [paper].

[8] Quan Zhang, Yongqiang Tian, **Yifeng Ding**, Shanshan Li, Chengnian Sun, Yu Jiang, and Jiaguang Sun. "CoopHance: Cooperative Enhancement for Robustness of Deep Learning Systems". *32nd ACM SIGSOFT International Symposium on Software Testing and Analysis (**ISSTA'23**)*. [paper].

[9] Quan Zhang, **Yifeng Ding**, Yongqiang Tian, Jianmin Guo, Min Yuan, and Yu Jiang. "AdvDoor: Adversarial Backdoor Attack of Deep Learning System". *30th ACM SIGSOFT International Symposium on Software Testing and Analysis (**ISSTA'21**)*. [paper].

Experiences

Since 05/2024 **Applied Scientist Intern, AWS AI Labs**
Hosted by: Hantian Ding, Zijian Wang
Developing novel pre-training paradigms [1] to advance code LLMs.

Since 04/2024 **Member, BigCode Project**
Contributed to StarCoder2-Instruct [3], where I instruction-tuned StarCoder2.

Since 08/2022 **Research Assistant, University of Illinois Urbana Champaign**
Advised by: Lingming Zhang
Building and enhancing LLMs for code generation [2, 3, 4] and program repair [7].

Academic Services

Reviewer AISTATS'25, ICLR'25, NeurIPS'24, ACL'24, EMNLP'24, NAACL'25

Organizing Committee LLM4Code'25 (International Workshop on Large Language Models for Code, co-organized with ICSE'25), LLM4Code'24

Invited Talks

11/2024 CAMEL-AI.org: "Horizon-Length Prediction: Advancing Fill-in-the-Middle Capabilities for Code Generation with Lookahead Planning"

11/2024 UCLA AGI Lab: "Improving Code Language Modeling via Horizon-Length Prediction"

10/2024 UIUC FM/SE Seminar: "Horizon-Length Prediction: Advancing Fill-in-the-Middle Capabilities for Code Generation with Lookahead Planning"

07/2024 Amazon Comprehend Team: "XFT: Unlocking the Power of Code Instruction Tuning by Simply Merging Upcycled Mixture-of-Experts"

04/2023 Uber Programming Systems Team: "Equipping Large Language Models with Domain-Specific Knowledge for Automated Program Repair"

Honors & Awards

Since 2022 **CS PhD Fellowship**, University of Illinois Urbana-Champaign

2021 **Research Excellence Scholarship**, Tsinghua University

2019 **Academic Excellence Scholarship**, Tsinghua University